A route map to calibrate spatial interaction models from GPS movement data

K. Sila-Nowicka[1], A.S. Fotheringham[2]

[1]Urban Big Data Centre
School of Political and Social Sciences
University of Glasgow
Lilybank Gardens 7, Glasgow, UK
G12 8RZ
Email: Katarzyna.sila-nowicka@glasgow.ac.uk

[2]School of Geographical Sciences and Urban Planning
Arizona State University
Coor Hall, 5th floor
975 S. Myrtle Ave.
Tempe, AZ 85287

1. Introduction

Human mobility is important for understanding the development and form of urban areas, the spatial distribution of facilities, and the provision of transportation services. Until recently, exploring human mobility in detail was challenging because personal trip data collection methods consisted of time-consuming and cost-intensive paper-and-pencil interviews, computer-assisted telephone interviews or computer-assisted self-interviews (Wolf et al., 2001). The development of sensors such as GPS trackers that capture movement data in real-time and at detailed spatial and temporal scales has transformed our ability to collect mobility data (Kwan and Neutens, 2014). Currently, however, these developments in movement data collection technologies are much further ahead than current methods for extracting meaningful patterns from such data (Laube et al., 2007; Long and Nelson, 2012). Furthermore, there have been very few studies that have tried to explore decision-making processes related with movements using GPS data and similar sources. There is a need therefore to discover if new forms of mobility data can be translated into new insights into movement behaviour. Here we do this through an examination of the calibration of spatial interaction with GPS data. As part of this experiment, we discover what needs to be done to transform raw GPS data into origin-destination matrices of aggregated flows.

2. Methods and applications

2.1 Data and case study

In order to investigate the usability of GPS movement data to calibrate spatial interaction models, we designed a GPS-based travel survey where participants were asked to carry GPS trackers continuously for a period of seven consecutive days (more in Sila-Nowicka et al. 2015). The collected dataset consisted of the traces from 206 individuals comprising 3,869,831 raw GPS locations (sampling rate 5 seconds), with each location record containing participant ID, latitude, longitude, elevation, date and time. Figure 1 shows the spatial extent of the collected trajectories.

To extract information from the GPS movement data, we filtered, pre-processed, segmented (using Spatio-Temporal Kernel Window), classified (using two-step feedforward neural network with general backpropagation algorithm) and contextually enriched these data using a framework for mobility patterns analysis (travel mode and places) from a combination of GPS movement data and contextual information (more information about the methods and the results can be found in Sila-Nowicka et al., 2015).

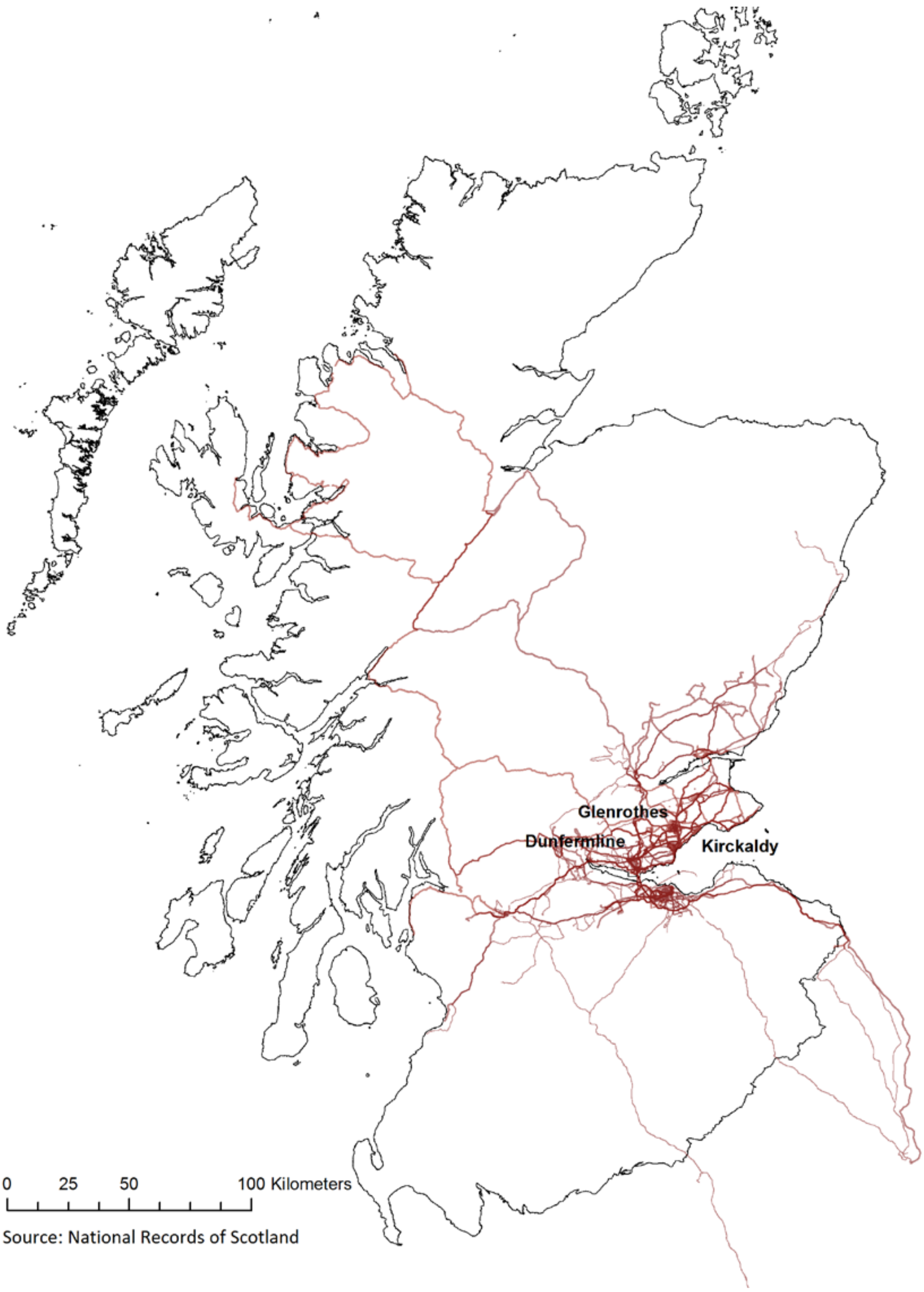


Figure 1: The extent of GPS movement data collected in the three largest towns (Dunfermline, Glenrothes and Kirkcaldy) in the Kingdom of Fife, UK.

2.2 Origin-destination matrices

With the segmented and classified GPS trajectories, the trip chains for each of the individuals were created. Using the trip chain structures, we were able to extract individual-thematic trips, such as commuting trips, shopping or leisure trips.

Table1. An example of the trip chain derived from the GPS trajectory.

User_id	Start tstamp	Stop tstamp	Time spent [seconds]	Mode/purpose	Geographic unit
8	06:22	06:39	995	home	Datazone A
8	06:58	15:25	30421	work	Datazone B
8	15:42	15:55	780	shopping	Datazone C
8	16:12	17:55	6169	home	Datazone A
8	17:55	18:32	2236	walk	Datazone A
8	18:32	20:21	6551	home	Datazone A
8	20:25	20:29	210	shopping	Datazone B
8	20:34	21:26	3133	home	Datazone A

Places of residence or work were aggregated to small administrative units (datazones) to offset privacy issues. By combining trips to common origins and destinations the data were encoded in an origin-destination matrix and used as input to calibrate (using MLE calibration process) maximum entropy spatial interaction models. Table 1 shows an example of the derived trip chains for one user, where the purpose of the journey and the travel mode to the destination could be provided.

2.3 Commuting and shopping behaviour – spatial interaction models

The goal of this paper is to provide evidence for the usability of GPS data to calibrate spatial interaction models and answer the question: "what insights can be gained into mobility decision-making through the use of such data?". To calibrate the models, we used origin-destination matrices derived from the GPS data of individual commuting flows and home-based shopping trips. To investigate the representativeness of the GPS data, we also calibrated models using official flow data (also referred to as interaction data that relate to the movement of people between places) from the 2011 Census (CIDER - Centre of Interaction Data Estimation and Research). In comparison to the CIDER data which are taken over a single time period, GPS-derived spatial interaction data provide the possibility of calibrating models on a weekly, daily or even hourly basis. In order to show that it is possible to use GPS-derived data to calibrate spatial interaction models, we took all the commuting flows for a week-long period within the study area of Dunfermline and calibrated a doubly constrained model. The predictive power of the GPS-based model is lower (R^2=0.5) than CIDER-based model (R^2=0.7) due to the huge difference in the number of flows collected. Furthermore, a much stronger distance effect for flows is recorded using the GPS data (β_{GPS}=-1.763; β_{CIDER}=-1.013). Visualisations of the results for Dunfermline are presented in Figure 2.

Next, we calibrated an origin-constrained model to investigate shopping behaviour in Dunfermline in the Kingdom of Fife, UK. The model is calibrated using only home-based shopping trips derived from the trip chains presented in Table 1.

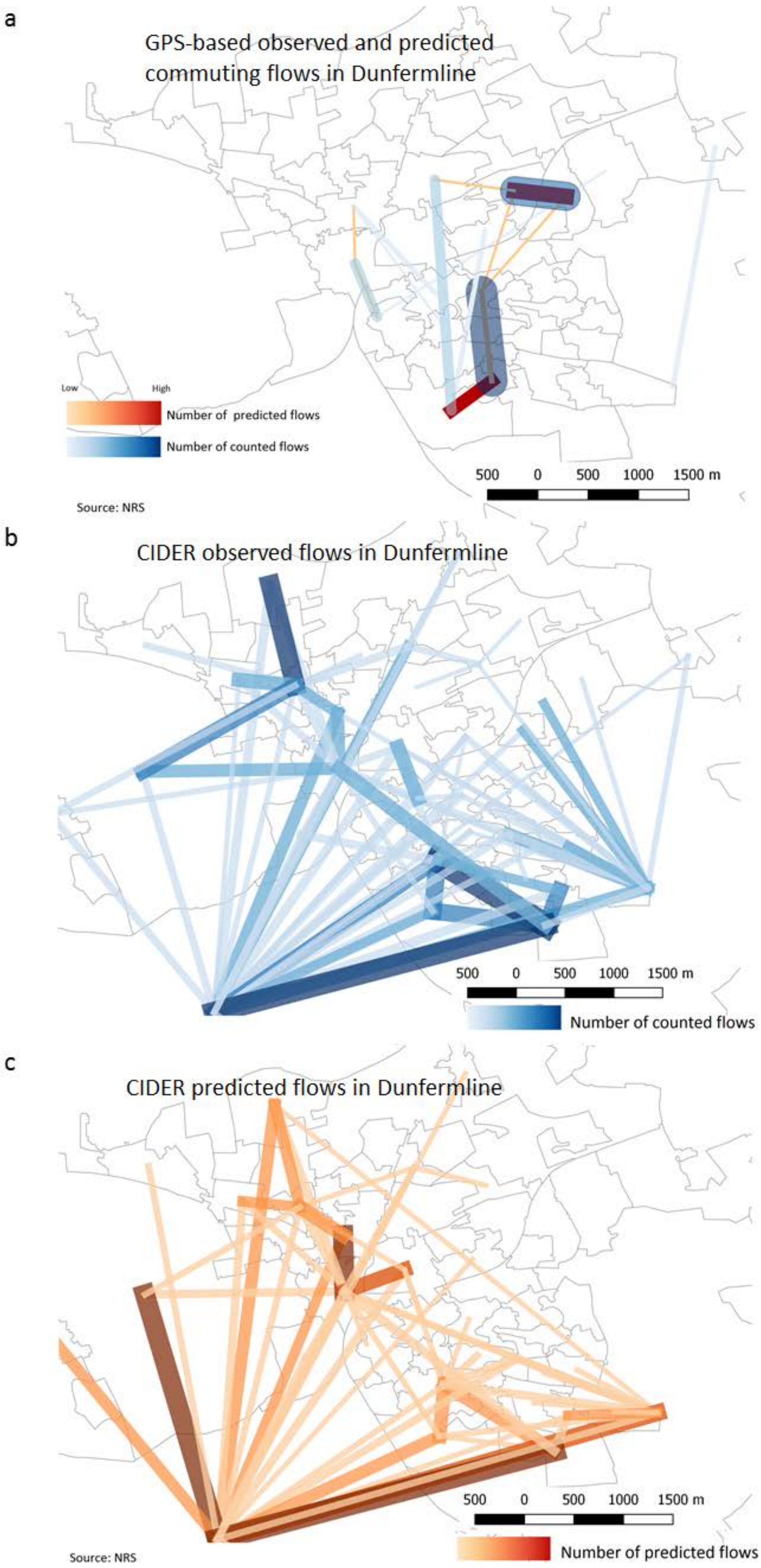
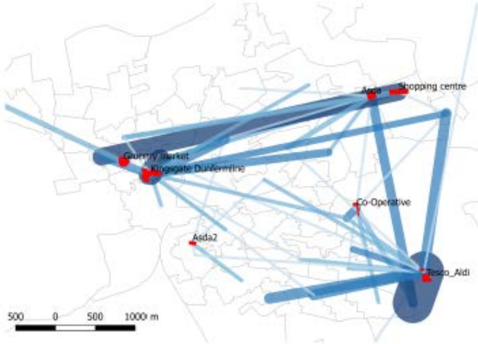
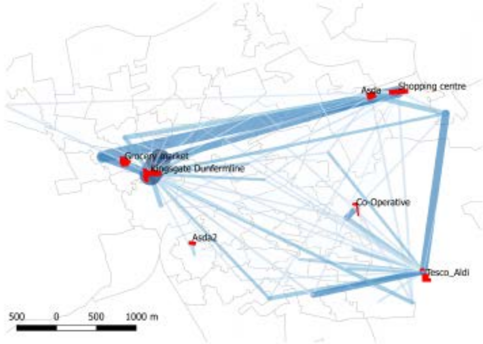


Figure 2: Observed and predicted flows for GPS movement data and Census-CIDER data.

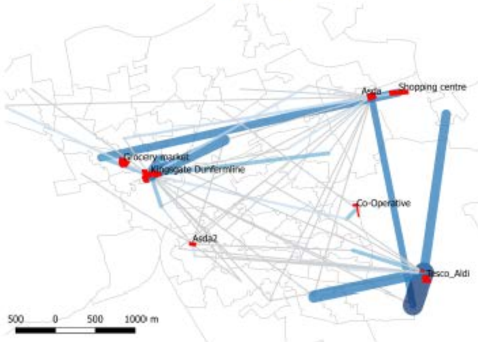
In order to obtain the origin-destination matrices to proceed with the model calibration, we first processed the GPS data to extract shopping locations from the GPS traces of the survey respondents.



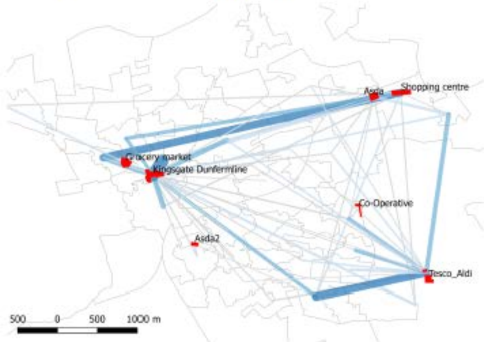
(a) Total observed shopping flows in Dunfermline



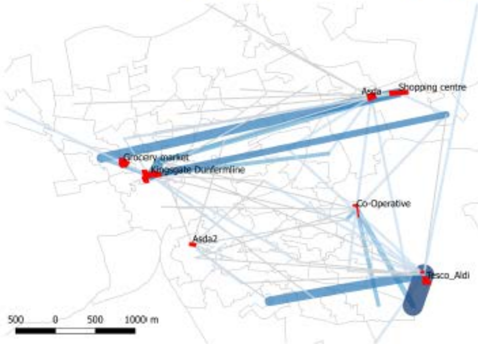
(b) Total predicted shopping flows in Dunfermline



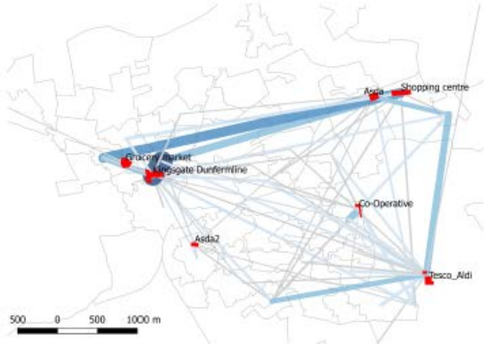
(c) Observed shopping flows on rainy days



(d) Predicted shopping flows on rainy days



(e) Observed shopping flows on the dry days



(f) Predicted shopping flows on the dry days

Figure 3: Observed and predicted shopping flows for GPS movement data.

The origin-constrained spatial interaction model requires an attractiveness variable for retail areas to be defined. In this case we used the area of each shopping unit. The results presented in Figure 3a and Figure 3b show the observed and predicted shopping flows in Dunfermline. The calibrated model underestimates the number of predicted flows for smaller retail areas. As expected, the size of the store has a positive impact on the choice of the store but this is relatively small (α=0.614) compared to some other retail studies where parameters around 1.0 and higher have been reported.. The distance decay parameter (β = -1.023) suggests that distance is a deterrent to shopping with a value close to many found in retail studies. Distance appears to have a near-linear effect on shopping propensity. The R2 of the model is equal to 0.776.

Because GPS data are used, we are able to calibrate separate models for various temporal intervals: every day, every weekday or weekend and therefore analyse shopping trips and compare temporal differences in shopping behaviour. The effect of weather on consumer behaviour and spending has received only limited attention in the marketing literature. Therefore, we designed a methodology to assign weather conditions to the GPS data points. Having done this we are able to calibrate weather-specific spatial interaction models. The results of weather-specific origin-constrained model show that size of the store has a greater positive impact on the choice of the store when it rains (α_{Rain}=0.805, α_{Dry}=0.494), the estimated distance decay parameters suggest that rain has limited affect on store choice (β_{Rain}= -0.972, β_{Dry}= -0.930). The R2 is equal to 0.690 for the rainy conditions and 0.746 for the model calibrated using only dry trips. The results for the weather-specific models are shown as visualisations in Figure 3c to Figure 3f. To our knowledge, time-specific and weather-specific GPS-based spatial interaction models have not been studied before.

3. Discussion

We live in an era in which vast new sources of mobility data are emerging that will allow spatial behaviour to be modelled by time of day, day of week and under different weather conditions. This study has shown this is possible. It has provided a route map to go from raw GPS data to calibrated spatial interaction models. It is inevitable that these new forms of data will replace standard types of survey and census data and therefore it is important that a robust framework for interpreting the data is presented.

Although we highlight the potential of GPS traces for the identification of human mobility patterns, as a recommendation for further studies we would suggest the collection of additional socio-demographic information in conjunction with GPS tracked locations to better ground truth trip modes and purposes.

Acknowledgements

The authors would like to thank all the survey participants. The research leading to these results has received funding from the European Union Seventh Framework Programme - Marie Curie Actions, Initial Training Network GEOCROWD under grant agreement n° FP7-PEOPLE-2010-ITN-.

References

Fotheringham, A.S. and O’Kelly, M.E., 1989, Spatial Interaction Model: Formulations and Applications. Kluwer Academic Publishers, London

Kwan, M.P. and Neutens, T., 2014, Space-time research in GIScience, International Journal of Geographical Information Science 28(5), 851-854.

Laube, P., Dennis, T., Forer, P. and Walker, M., 2007, Movement beyond the snapshot - Dynamic analysis of geospatial lifelines, Computers, Environment and Urban Systems 31(5), 481-501.

Long, J. A. and Nelson, T. A. 2012, `A review of quantitative methods for movement data', International Journal of Geographical Information Science 27(2), 1-27.

Sila-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demsar, U. and Fotheringham, A. S., 2015, Analysis of human mobility patterns from GPS trajectories and contextual information, International Journal of Geographical Information Science pp. 1-26.

Wolf, J., Guensler, R. and Bachman, W., 2001, Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data, Transportation Research Record 1768(1), 125-134.